



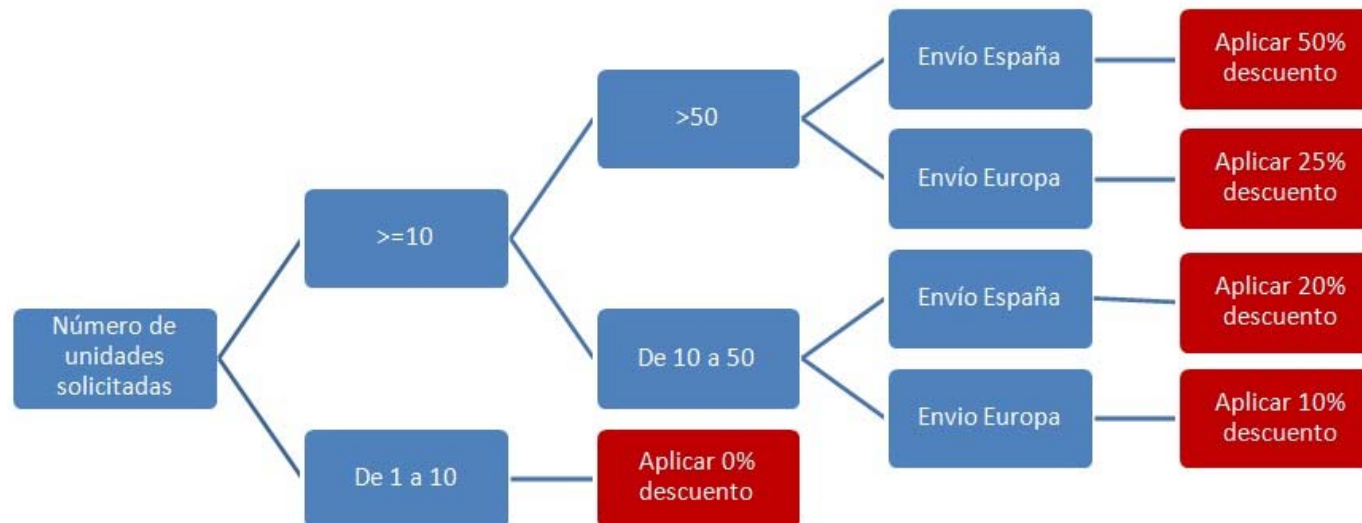
POLITÉCNICA

Proyecto 6



□ Árboles de decisión:

- Un árbol de decisión es un **modelo de predicción** utilizado en el ámbito de la inteligencia artificial.





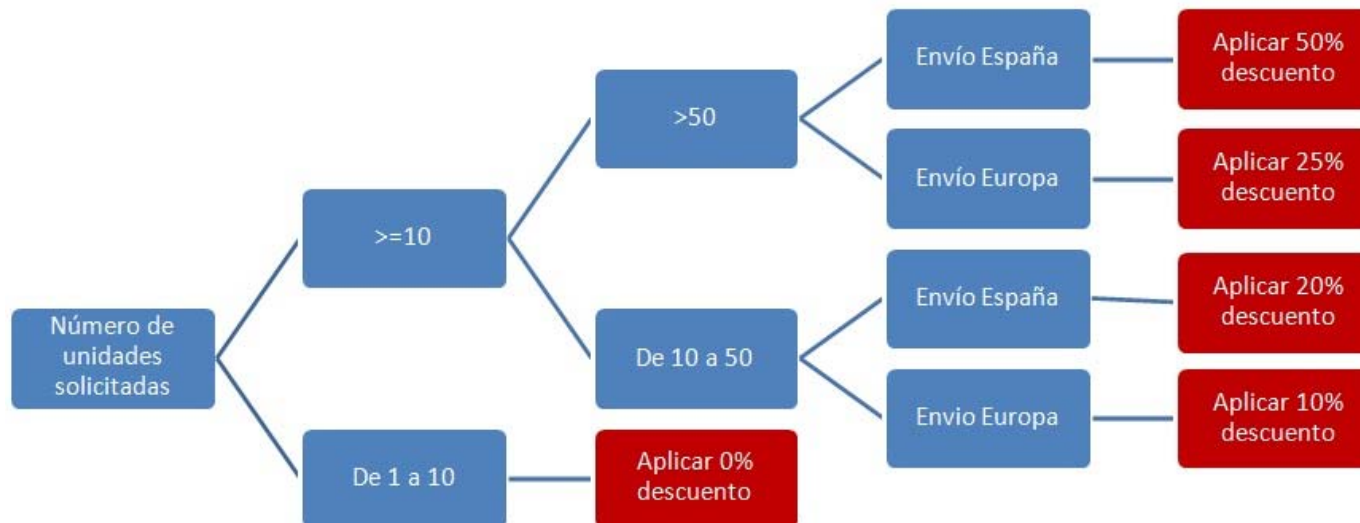
POLITÉCNICA

Proyecto 6



❑ Funcionamiento:

- Se realiza un **test en cada nodo interno** del árbol, a medida que se recorre desde la raíz hasta las hojas, para **alcanzar así una decisión**.





POLITÉCNICA

Proyecto 6



❑ Características:

- Un árbol de decisión indica las **acciones a realizar** en función del valor de una o varias variables (nodos internos) hasta llegar a una decisión (hoja).
- Es una **representación en forma de árbol** cuyas ramas se bifurcan en función de los valores tomados por las variables y que terminan en una acción concreta.
- Se suele utilizar cuando el **número de condiciones no es muy grande**. En tal caso, es mejor utilizar una tabla de decisión.
- Si utilizamos valores discretos el árbol de decisión se suele llamar **clasificación**, mientras que si los valores son continuos se llama **regresión**.



POLITÉCNICA

Proyecto 6



□ Algoritmo ID3:

- El algoritmo ID3 es utilizado dentro del ámbito de la **inteligencia artificial**.
- Su uso se engloba en la **búsqueda de hipótesis o reglas**, dado un conjunto de ejemplos (*Machine learning*).
- El algoritmo ID3 **construye árboles de decisión**.
- El conjunto de ejemplos deberá estar conformado por una serie de tuplas de valores (**atributos**), en el que uno de ellos (el atributo a clasificar) es el **objetivo** y es **de tipo binario**.



POLITÉCNICA

Proyecto 6



□ Un conjunto de *ejemplos*:

Día	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14
Cielo	S	S	N	Ll	Ll	Ll	N	S	S	Ll	S	N	N	Ll
Temp.	A	A	A	S	B	B	B	S	B	S	S	S	A	S
Hum.	A	A	A	A	N	N	N	A	N	N	N	A	N	A
Viento	D	F	D	D	D	F	F	D	D	D	F	F	D	F
Jugar	-	-	+	+	+	-	+	-	+	+	+	+	+	-

- **Cielo:** Sol (S), Nubes (N), Lluvia (Ll)
- **Temperatura:** Alta (A), Sueve (S), Baja (B)
- **Humedad:** Alta (A), Normal (N)
- **Viento:** Fuerte (F), Débil (D)



POLITÉCNICA

Proyecto 6



□ ¿En qué orden consideramos los atributos?

Día	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14
Cielo	S	S	N	LI	LI	LI	N	S	S	LI	S	N	N	LI
Temp.	A	A	A	S	B	B	B	S	B	S	S	S	A	S
Hum.	A	A	A	A	N	N	N	A	N	N	N	A	N	A
Viento	D	F	D	D	D	F	F	D	D	D	F	F	D	F
Jugar	-	-	+	+	+	-	+	-	+	+	+	+	+	-

▪ Entropía de un conjunto:

$$-\left(\frac{p}{d}\right) \log_2\left(\frac{p}{d}\right) - \left(\frac{n}{d}\right) \log_2\left(\frac{n}{d}\right)$$

- p : número de valores +
- n : número de valores -
- d : número total de elementos



Proyecto 6

□ ¿En qué orden consideramos los atributos?

Día	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14
Cielo	S	S	N	LI	LI	LI	N	S	S	LI	S	N	N	LI
Temp.	A	A	A	S	B	B	B	S	B	S	S	S	A	S
Hum.	A	A	A	A	N	N	N	A	N	N	N	A	N	A
Viento	D	F	D	D	D	F	F	D	D	D	F	F	D	F
Jugar	-	-	+	+	+	-	+	-	+	+	+	+	+	-

▪ Ganancia de información de un atributo A:

$$IG(A, S) = H(S) - \sum_{t \in T} p(t)H(t)$$

- t = subconjunto para un valor dado de A
- T = conjunto de subconjuntos que definen los valores de A



Proyecto 6

□ ¿En qué orden consideramos los atributos?

Día	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14
Cielo	S	S	N	LI	LI	LI	N	S	S	LI	S	N	N	LI
Temp.	A	A	A	S	B	B	B	S	B	S	S	S	A	S
Hum.	A	A	A	A	N	N	N	A	N	N	N	A	N	A
Viento	D	F	D	D	D	F	F	D	D	D	F	F	D	F
Jugar	-	-	+	+	+	-	+	-	+	+	+	+	+	-

▪ Ganancia de información del atributo *Cielo*:

- $t_S = \{D1, D2, D8, D9, D11\}$ ▪ $p(t_S) = 5/14$ ▪ $H(t_S) = 0,97$ 2-3
- $t_N = \{D3, D7, D12, D13\}$ ▪ $p(t_N) = 4/14$ ▪ $H(t_N) = 0$ 4-0
- $t_{LI} = \{D4, D5, D6, D10, D14\}$ ▪ $p(t_{LI}) = 5/14$ ▪ $H(t_{LI}) = 0,97$ 3-2



Proyecto 6

□ ¿En qué orden consideramos los atributos?

Día	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14
Cielo	S	S	N	LI	LI	LI	N	S	S	LI	S	N	N	LI
Temp.	A	A	A	S	B	B	B	S	B	S	S	S	A	S
Hum.	A	A	A	A	N	N	N	A	N	N	N	A	N	A
Viento	D	F	D	D	D	F	F	D	D	D	F	F	D	F
Jugar	-	-	+	+	+	-	+	-	+	+	+	+	+	-

▪ Ganancia de información del atributo *Cielo*: $H(S) = 0,94$ 9-5

- $p(t_S) = 5/14$ ▪ $H(t_S) = 0,97$
- $p(t_N) = 4/14$ ▪ $H(t_N) = 0$
- $p(t_{LI}) = 5/14$ ▪ $H(t_{LI}) = 0,97$

$IG(Cielo, S) = 0,25$



POLITÉCNICA

Proyecto 6



□ ¿En qué orden consideramos los atributos?

Día	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14
Cielo	S	S	N	LI	LI	LI	N	S	S	LI	S	N	N	LI
Temp.	A	A	A	S	B	B	B	S	B	S	S	S	A	S
Hum.	A	A	A	A	N	N	N	A	N	N	N	A	N	A
Viento	D	F	D	D	D	F	F	D	D	D	F	F	D	F
Jugar	-	-	+	+	+	-	+	-	+	+	+	+	+	-

- Se elige el atributo que da mayor ganancia de información



POLITÉCNICA

Proyecto 6



□ Algoritmo ID3 (recursivo)

Id3(Ejemplos, Objetivo, Atributos)

- Si todos los ejemplos son positivos devolver un nodo positivo
- Si todos los ejemplos son negativos devolver un nodo negativo
- Si Atributos está vacío devolver un nodo con el voto mayoritario del valor del Objetivo en Ejemplos
- En otro caso:
 - Sea A el MEJOR de Atributos
 - Para cada valor v del atributo A hacer:
 - Sea Ejemplos(v) el subconjunto de Ejemplos con valor de A v
 - Si Ejemplos(v) está vacío devolver un nodo con el voto mayoritario del Objetivo de Ejemplos
 - Sino devolver Id3(Ejemplos(v), Objetivo, Atributos- $\{A\}$)



POLITÉCNICA

Proyecto 6



□ Input: archivo con el siguiente formato

m <número de atributos (el último es el objetivo)>
t <tipo del primer atributo (t = s → string ó t = i → entero)>
Si el t anterior es s:
c <primer valor posible del primer atributo>
c <segundo valor posible del primer atributo>
En otro caso (si t es i):
k
...
b <tipo del último atributo (debe ser obligatoriamente b)>
n <número de ejemplos>
v <valor del primer atributo del primer ejemplo>
...
v <valor del último atributo del primer ejemplo>
...
v <valor del primer atributo del último ejemplo>
...
v <valor del primer último del último ejemplo>

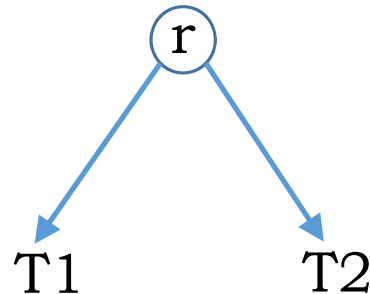


POLITÉCNICA

Proyecto 5



□ Output: cadena representando el árbol de decisión



Si el atributo de la raíz r es una cadena:

$Salida(T)=[Atributo(r),v1(r),v2(r),Salida(T1),Salida(T2)]$

Si el atributo de la raíz r es un entero:

$Salida(T)=[Atributo(r),v(r),Salida(T1),Salida(T2)]$